QUALITY OF SERVICES (QoS) ANALYSIS FRAMEWORK FOR TEXT TO SPEECH (TTS) SERVICES

MOHAMUD MOHAMED HASSAN

A dissertation submitted in partial
fulfillment of the requirements for the award of the
Degree of Master of Computer Science (Web Technology)

Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia

SEPTEMBER 2015

This thesis is dedicated for my beloved parents,

Mohamed Hassan, Halwo Elmi and Mulaho Hassan.

ACKNOWLEDGEMENT

First and foremost, all praise are due to Allah (the creator), who helped and gave me the strength, patience and knowledge to continue and finish my studies. Secondly, I would like to extend my heartfelt gratitude to my supervisor, Dr. Mohd Farhan MD Fudzee for his invaluable guidance, helpful suggestions and motivations during my studies and throughout this dissertation.

I would like to thank all my lectures and the staff of the Faculty Computer Science and Information Technology.

Lastly, I extend my acknowledgment and heartfelt love to my family for their financially and moral support throughout my study with all their love, encouragement.

ABSTRACT

Quality of service (QoS) evaluations is significant and necessary for text to speech web service applications. Text to speech media conversion quality measurements has general and specific mechanisms for its functional and nonfunctional requirements. The main objective of this thesis is to introduce QoS framework which is able to evaluate and analyze the perceived quality of services (QoS) for text to speech (TTS) web services. To achieve this goal, the framework combines two main mechanisms for measuring the speech quality. General quality attributes measure the response time of TTS services, specific quality attributes measure intelligibility and naturalness through subjective quality measurements, which are mapped onto mean opinion score (MOS). Twenty individuals participated the experiment to test the speech quality by comparing three services fromtexttospeech.com, Natural Reader and Yakitome. Aggregate scores has been used to calculate the combination of general and specific nonfunctional QoS on TTS Web services. The result shown better scale for quality estimation, service1 (Fromtexttospeech) 47.84% is suitable TTS service provider where service2 and service3 (NaturalReader and Yakitome) are close 31.62 and 21.53% respectively and less preferred for listening tests to assess synthesized speech. It is essential to consider the user's perspective when evaluating the quality of services for media conversion services such as text to speech (TTS) to enhance the user experience.

ABSTRAK

Penilaian kualiti perkhidmatan (QoS) adalah penting dan perlu bagi aplikasi perkhidmatan web teks-ke-tutur. Ukuran kualiti penukaran media teks–ke-tutur mempunyai mekanisme umum dan khusus bagi keperluan fungsian dan bukan fungsian. Objektif utama projek ini adalah untuk memperkenalkan rangka kerja QoS yang mampu menilai dan menganalisis tanggapan kualiti perkhidmatan (QoS) untuk perkhidmatan web teks–ke–tutur (TTS). Untuk mencapai matlamat ini, rangka kerja tersebut menggabungkan dua mekanisme utama untuk mengukur kualiti pertuturan. Sifat-sifat kualiti umum mengukur tempoh tindak balas perkhidmatan TTS, sifat-sifat kualiti khusus mengukur kebolehfahaman dan kebersahajaan melalui ukuran kualiti subjektif, yang dipetakan ke min skor pendapat (MOS). Dua puluh individu menyertai eksperimen ini untuk menguji kualiti pertuturan dengan membandingkan tiga perkhidmatan (fromtexttospeech.com, Natural Reader dan Yakitome). Untuk mengira gabungan bukan fungsian umum dan khusus QoS di perkhidmatan Web TTS, kami menggunakan skor agregat. Hasilnya menunjukkan skala yang lebih baik untuk anggaran kualiti, servis 1 (Fromtexttospeech) 47.84% adalah pembekal perkhidmatan TTS yang sesuai di mana servis 2 dan servis 3 (NaturalReader dan Yakitome) adalah hampir dengan masing-masing 31.62 dan 21.53% dan kurang menjadi pilihan untuk ujian pendengaran bagi menilai pertuturan yang disintesis. Penilaian kualiti perkhidmatan untuk perkhidmatan penukaran media seperti perkhidmatan teks–ke-tutur (TTS) adalah penting untuk mempertimbangkan perspektif pengguna yang boleh menyediakan pengguna dengan keupayaan untuk meningkatkan pengalaman pengguna.

TABLE OF CONTENTS

TITLE		i
DECLARATION		ii
DEDICATION		iii
ACKNOWLEDGEMENT		iv
ABSTRACT		v
ABSTRAK		vi
TABLE OF CONTENTS		vii
LIST OF FIGURES		x
LIST OF TABLES		xi
CHAPTER 1 INTRODUCTION		
1.1	Background	1
1.2	Problem Statement	3
1.3	Research Objectives	4
1.4	Scope of the Research	4
1.5	Significance of Research	4
CHAPTER 2 LITERATURE REVIEW		
2.1	Introduction	5
2.2	Architecture of Text-to-speech systems	6
	2.2.1 Text Analysis and Detection	7
	2.2.2 Text Normalization and Linearization	7
	2.2.3 Phonetic Analysis	7
	2.2.4 Prosodic Modeling and Intonation	8
	2.2.5 Acoustic Processing	9
2.3	Requirements of QoS attributes	9
	2.3.1 Performance Requirements	10
	2.3.2 Nonfunctional Quality Attributes	11

2.3.2.1 Intelligibility and Reading comprehension 12
2.3.2.2 Naturalness 12
2.4 Determination of offered QoS 13
2.5 Techniques used to evaluate the QoS of TTS 14
2.5.1 Mean Opinion Score (MOS) 15
2.5.2 Semantically Unpredictable Sentences (SUS) 17
2.6 Related work 16
2.7 QoS analysis for web based TTS 18
2.8 Summary 18
CHAPTER 3 METHODOLOGY
3.1 Introduction 19
3.2 Requirement of QoS and QoE 19
3.3 Proposed Framework 20
3.3.1 Evaluating Quality of Service (QoS) attributes 22
3.3.2 Combine Method 23
3.3.3 Decision making process 23
3.4 Designing the Framework 23
3.5 QoS measurement Algorithm 24
3.6 Mean Opinion Scale (MOS) Scale 25
3.7 Summary 26
CHAPTER 4 IMPLEMENTATION
4.1 Introduction 27
4.2 TTS Speech analysis 27
4.2.1 Capturing TTS Speech information 28
4.2.2 Calculating response time 30
4.3 Evaluations for intelligibility and naturalness 31
4.3.1 MOS Design and structure 32
4.3.2 MOS Intelligibility 32
4.3.3 MOS Naturalness 34
4.3.4 Database Design for MOS 36

4.4 MOS Data presentation 37
4.5 Combined service calculation 39
4.6 UI data presentation 40
4.7 Summary 42

CHAPTER 5 RESULTS AND DISCUSSIONS
5.1 Introduction 43
5.2 Participants 43
5.2.1 TTS Speech Test 44
5.3 MOS analysis result 44
5.3.1 MOS Intelligibility result analysis 45
5.3.2 MOS Naturalness result analysis 50
5.4 Overall Quality analysis 55
5.5 Summary 57
CHAPTER 6 CONCLUSION AND FUTURE WORK
6.1 Introduction 58
6.2 Contribution of the research 59
6.3 Suggestions for future work 59
REFERENCES 60
APPENDICES 65

LIST OF FIGURES

1.1	Process of text-to-speech (TTS) web system	2
2.1	Architecture of Text-to-speech systems	6
2.2	Text to speech (TTS) nonfunctional Quality attributes	13
2.3	Process for the determination of offered Quality of services	14
3.1	Framework for QoS analysis for TTS services	21
3.2	Flowchart for measuring the QoS for TTS services	24
4.1	JSON code for capturing information of the TTS speech file.	29
4.2	Algorithm processes for calculating response time.	31
4.3	PHP code for intelligibility data analysis from database and pass to JSON script.	33
4.4	PHP code for naturalness data analysis from database and pass to JSON script.	35
4.5	MOS database storage including the relationships.	36
4.6	MOS questions for intelligibility and naturalness.	38
4.7	Shows the process of user interface (UI) for QoS analysis for text to speech.	41
5.1	MOS intelligibility performance comparison for online TTS services.	49
5.2	MOS naturalness performance comparison for online TTS services.	54
5.3	Overall quality of service analysis for online TTS services.	56

LIST OF TABLES

4.1	List of TTS services that will be used to examine the QoS analysis.	28
4.2	Quality of services (QoS) categorization.	39
5.1	Shows the calculation result for response time.	44
5.2	MOS intelligibility quality performance for fromtexttospeech TTS services (by student participants)	45
5.3	MOS intelligibility quality performance for fromtexttospeech TTS services (by non-student participants)	46
5.4	MOS intelligibility quality performance for NaturalReader TTS services (by student participants)	46
5.5	MOS intelligibility quality performance for NaturalReader TTS services (by non-student participants)	47
5.6	MOS intelligibility quality performance for Yakitome TTS services (by student participants)	47
5.7	MOS intelligibility quality performance for Yakitome TTS services (by non-student participants)	48
5.8	MOS intelligibility performance for average quality of compared online TTS services.	48
5.9	MOS naturalness performance for Fromtexttospeech online TTS services (bystudent participants)	50
5.10	MOS naturalness performance for Fromtexttospeech online TTS services (by non-student participants)	51
5.11	MOS naturalness performance for NaturalReader online TTS services (bystudent participants)	51

5.12	MOS naturalness performance for NaturalReader online TTS services (by non-student participants)	52
5.13	MOS naturalness performance for Yakitome online TTS services (by student participants)	52
5.14	MOS naturalness performance for Yakitome online TTS services (by non-student participants)	53
5.15	MOS naturalness performance for online TTS services	53
5.16	QoS performance with the computed aggregate score	55

CHAPTER 1

INTRODUCTION

1.1 Introduction

Text to speech systems (TTS) are important in our daily activities for work, education, communication and many more. The use of speech synthesis and in building voices became common due to the rapid advancement in information technology and communications. Ultimately, high quality synthesized outputs are preferred. Thus evaluating the TTS web services are impotent. Text to speech (TTS) is useful in the areas like disabled, education, consumer, computer interface and telecommunications. There are many TTS web based applications available. These applications provide different services depending on their target demographic audience and are usually engineered to suite specific quality of services [1, 2, 3]. The voice user interface (VUI) plays huge role in technology such as computer systems, mobile multimedia and voice-enabled equipment. Speech Understanding and Synthesis Technology are among the most frequently used technology to support users. Texts to speech (TTS) systems are among these applications for speech technologies, where TTS systems read text on the screen of the user device aloud using synthesized speech [4].

TTS systems translate digital text into audio/ speech. Text to speech (TTS) applications assist individual who experience dyslexia, reading challenges, or visual impairment as well as foreign language learning for enhancing their listening skills. Students can also use text to speech (TTS) for listening to course materials as revision for their study [5, 6]. Figure 1.1 below shows the process of text to speech for web applications.

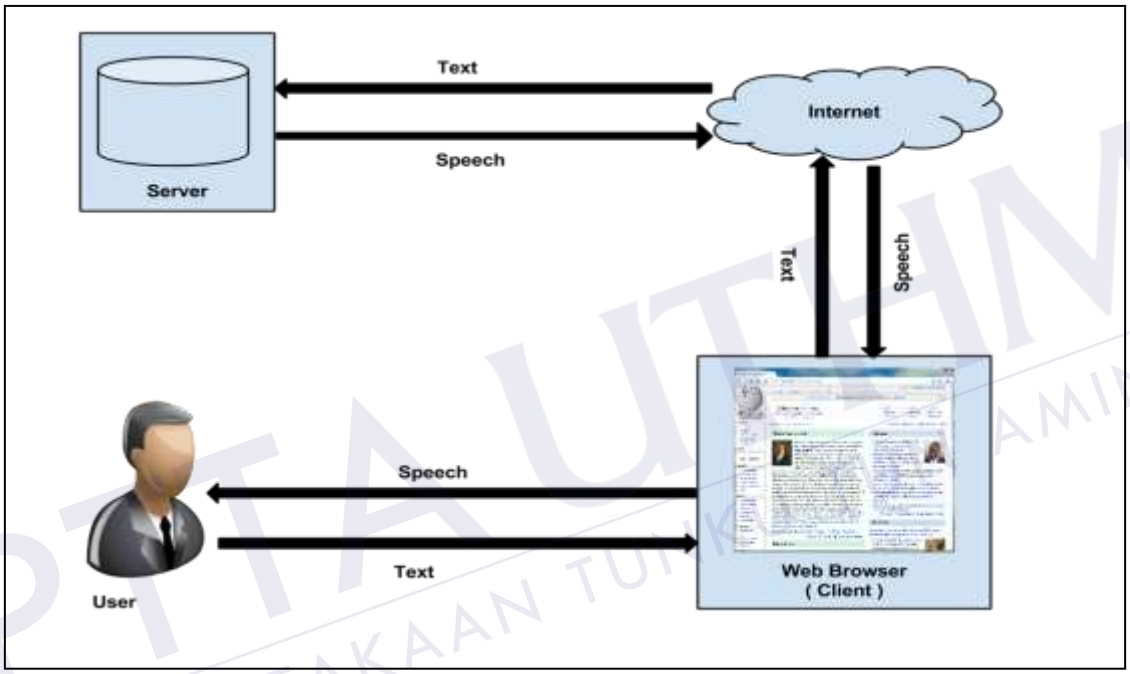

Figure 1.1 Process of text-to-speech (TTS) web system

TTS application developers use different technique to reach their client to provide and present their content in a suitable manner. Although, reaching wide range of client makes it more difficult to maintain the quality of the product. In the process of designing any software application, it's very important to consider the quality of service of the user experience. Normally, customers expect high quality from the service or the application that they use in different platforms. QoS performance parameters have different priorities for different user sectors.

The QoS in multimedia conversion of text to speech (TTS) examines the performance in terms of accessibility, media conversion availability, conversion accuracy, and user satisfactions. QoS determines how well a service performs while functionality determines what a service does [7].

1.2 Problem Statement

Text to speech (TTS) services quality evaluations can be classified into functional and nonfunctional requirements. Functional requirement focuses on what TTS service does, while nonfunctional requirements also known as Quality attributes is used to determine the quality of services requirements. There is mechanism for general QoS and there is mechanism for specific QoS which is used to evaluate nonfunctional requirements of TTS services. Nevertheless there is no effort to integrate both mechanisms into a single mechanism. There are currently many TTS services that users can access online. However, there are no QoS services evaluations that compare between TTS services. The existing QoS analyses for TTS are concentrated only server sides. By analyzing QoS from the end-user perspective it can provide users with capacity to enhance their experience.

The goal of this research is to identify and analyze quality of services (QoS) of text to speech in web services by integrating both mechanisms into single mechanism and presenting it to the end user to enhance the user experience.

The challenges that users face include:

i. How to discover text to speech applications with appealing quality of service?

ii. How to determine the qualities of services (QoS) of text to speech (TTS) web application that provides better services for the users?

1.3 Research Objectives

The objectives of this dissertation are to:

i. Identify and discover current quality of service for text to speech (TTS) multimedia conversion.
ii. Propose integrated quality of service (QoS) framework for text to speech multimedia conversion services.
iii. Implement the proposed quality of services (QoS) framework for TTS services.

1.4 Scope of the Research

This dissertation will examine the quality of services of the multimedia conversion text to speech on the web and will measure TTS performance in term of content accessibility, response time, and voice intelligibility and naturalness to enhance the quality of experience of the online users.

This research work will focus only on QoS for TTS media conversion on client side. This study will not cover information overload, data loss and content adjustment. Also it will not cover aspects of evaluating speech to text conversion as well.

1.5 Significance of Research

This research work is aimed on providing a contribution to the QoS analysis for text to speech (TTS) multimedia conversion web environment. By developing a framework that analyses the nonfunctional quality attributes of text to speech media conversion to help users to identify the quality of services (QoS) of text to speech (TTS) applications. The significance of this research is the capability to provide a feedback to the end users.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Text to speech system (TTS) is a production of artificial human voice from text to speech produced by means of computing device. In speech synthesis, the input is standard text or a phonetic spelling, and the output is a spoken version of the text. Speech synthesis allows generating and transforming strings of phonetics and prosodic symbols into a synthetic speech signal [8]. This transformation produces synthetic speech that is very close to real human voice in compliance with the communication norms of special languages. In the last couple of decades, TTS received a lot of attention from researchers to assist individual who experience dyslexia, reading challenges, or visual impairment as well as foreign language learning for enhancing listening skills [9].

2.2 Architecture of Text-to-speech systems

Text to speech systems involves an algorithmically converting process of an input text or selected text transformation into actual voice. The conversion is done by following various steps as seen in the figure 2.1 below.

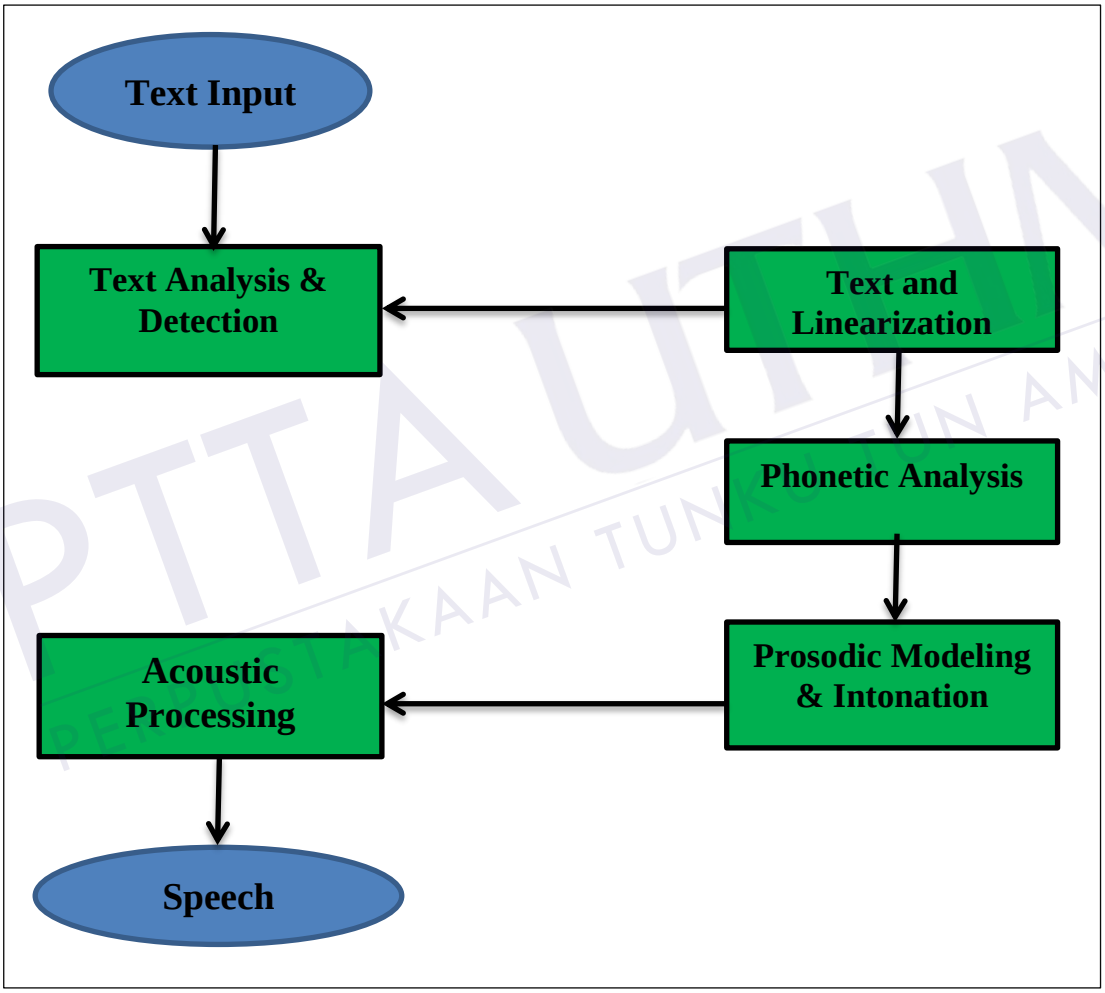

Figure 2.1 Architecture of Text-to-speech systems [8]

2.2.1 Text Analysis and Detection

Text analysis is first stage of text to speech conversion system, it will analyze an input text and organizes letters, numbers, abbreviations, acronyms and idiomatic and transforms them into manageable text. Text detection is capturing the text from different format of documents. For example extracting text from video, image of other documents like newspapers, books etc.

2.2.2 Text Normalization and Linearization

Text Normalization is the transformation of text to readable (pronounceable) form. Normalizing of text is normally performed before text is transformed to synthesized speech. This process is done through converting all letters of lowercase or upper case, to remove punctuations, accent marks, stopwords or "too common words" and other diacritics from letters. The main aim here is to identify punctuation marks and pauses between words. Text normalization is useful for comparing two sequences of characters which represented differently but mean the same. For example, "Don't" vs "Do not", "I'm" vs "I am", and many more [10].

2.2.3 Phonetic Analysis

Phonetic Analysis also known as grapheme-to-phoneme conversion, is a process of conversion the orthographical symbols into phonological ones using a phonetic alphabet. Phone is the smallest unit of sound, which is the sound that has definite shape as a sound wave. Phoneme is a collection of phones that constitute minimal distinctive phonetic units [11]. Speech synthesis systems use two basic approaches to determine the pronunciation of a word based on its spelling:

i. **Dictionary based approach:** a dictionary stores all different type of words with their correct pronunciation. This approach is very quick and accurate and the pronunciation quality will be better but the major drawback is that it needs a large database to store all words and the system will stop if a word is not found in the dictionary.

ii. **Rule based approach:** The letter sounds for a word are blended together to form a pronunciation based on some rule. Here, the main advantage is that it requires no database and it works on any type of input. same way the complexity grows for irregular inputs.

2.2.4 Prosodic Modeling and Intonation

The main objective of prosody is the combination of the stress pattern, rhythm of speech, and intonation in a speech. The prosodic modeling describes the emotions of the speakers, whereas Intonation is simply a variation of speech while speaking. Recent study suggested that identification of the vocal features that signal emotional content can be used to help make synthesized speech sound more natural. Modelling of an intonation is an important task that affects intelligibility and naturalness of the speech. To receive high quality text to speech conversion, good model of intonation is needed [12, 13].

Generally intonations are distinguished as:

i. Rising Intonation: when the pitch of the voice increases
ii. Falling Intonation: when the pitch of the voice decreases
iii. Dipping Intonation: when the pitch of the voice falls and then rises
iv. Peaking Intonation: when the pitch of the voice rises and then falls

2.2.5 Acoustic Processing

This phase speech signal will be produced according to voice characteristics needed to be read out laud. There are three types of Acoustics:

i. Concatenative Synthesis: prerecorded words of a person voice saved in to the database. The natural sounding speech is the main advantage and the main drawback is in the using and developing of large database.
ii. Formant Synthesis: this type of acoustic can be constantly intelligible, and it does not need any database of prerecorded speech samples. Therefore, formant synthesis produces artificial and robotic speech.
iii. Articulatory Synthesis: this type of acoustic uses techniques in which synthesizing speech is based on models of the human vocal tract are to be developed. Articulatory Synthesis produces a complete synthetic output, typically based on mathematical models [14, 15].

2.3 Requirements of QoS attributes

Requirements can be broadly classified into functional requirements and nonfunctional requirements. Functional requirements basically describe what the services do which has system functionality. Whereas, non-functional requirements deals generally with quality requirements which specify how well the services can perform. In text to speech media conversion nonfunctional requirement describes the design constraints, performance requirements, and software quality attributes. Performance requirements are including the accessibility, cost and response time of the TTS services and are called general quality attributes. Specific quality attributes: include intelligibility and naturalness of TTS services [16].

2.3.1 Performance Requirements

Performance requirements concern on the speed of text to speech services. There are numerous attributes that concerns performance requirements:

i. **Response time:** how quickly the system reacts to a user input. In text to speech (TTS) web response time denotes the time for each TTS page to load or response. The goal is to measure the accurate time for the required load time for TTS web performance.

ii. **Throughput requirements:** how much the system can accomplish within a specified amount of time. Text to speech (TTS) system must be capable of summing the speech file that is delivered to the web application users.

iii. **Accessibility requirements:** is the system accessible for service when requested by end-users. TTS services should be accessible through the internet.

2.3.1.1 Accessibility of TTS services

Accessibility metrics that produce quantitative scores enable accurate discrimination among web pages. Quantitative scores are useful in those scenarios where accurate measurement is required such as in web engineering, quality assurance, accessibility monitoring observatories and information retrieval. In recent years, a good deal of research has been dedicated to TTS web accessibility metrics. Existing metrics provide a general approach for measuring accessibility as they do not consider specific user groups but rather general purpose guideline-sets. While some are automatically obtained, others require human judgment [17].

2.3.1.2 Response Time

Text to speech systems (TTS) is developed in a way that services conversion is able to get real time immediate response even for a large sized text files since the TTS response time depends on the length of the size of the text. Response time measures the performance of text to speech systems (TTS) to reduce the delay time and to enhance processing time of the services [18, 19].

High quality speech synthesis is the ultimate goal where having accurate processing time is a very significant element that affects the quality of the synthesized speech. Response time gives online text to speech users experience on web speech by calculating the synthesized fundamental frequency contour. Synthesis speech has duration time which is most likely sequence of occurrences which match the estimated targets from a given text to processing.

Time scale prosodic adjustment algorithms utilize the predicted durations to calculate the time scale for the selected speech units. It is clear that process time contributes to the most of the components that are usually utilized during the process of converting text to speech. Therefore, accurate and robust process time is one of the fundamental and interesting tasks of developing TTS systems [20].

2.3.2 Nonfunctional Quality Attributes

Nonfunctional requirements are usually evaluated subjectively. To measure the Quality of services (QoS) in text to speech (TTS) it's very important to consider three main factors that drive text-to-speech TTS applications. Figure 2.2 shows nonfunctional quality attributes for text to speech (TTS) services.

2.3.2.1 Intelligibility and Reading comprehension

Intelligibility means how the listener / user can understand the spoken words and its units of all lengths, from phonemes to sentences. The smallest speech units (phonemes / syllables) are called segment units or segment testing. This type of measurement is calculated by the basis of amounts of correctly heard segments. While, reading comprehension is used to measure overall understanding of the speech in matter of language. The word "intelligibility" means the degree of each word being produced in a sentence; while the word "comprehension" means the degree of received messages being understood [21].

The intelligibility task tends to make listeners difficult to predict the unheard information, by listening to each sentence not more than one time, or to keep as few times as possible. The intelligibility ratings related to both phonological and literacy outcomes [22].

2.3.2.2 Naturalness

Naturalness is a highly important feature of synthetic speech. Apart from the segmental quality and the voice characteristics, it depends mostly from the prosody. Naturalness can be defined as how much the synthetic speech is similar to the natural speech. It also measures the overall quality of synthetic speech are abstract subjective attributes [23, 24]. Figure 2.2 shows nonfunctional Quality attributes for Text to speech (TTS).

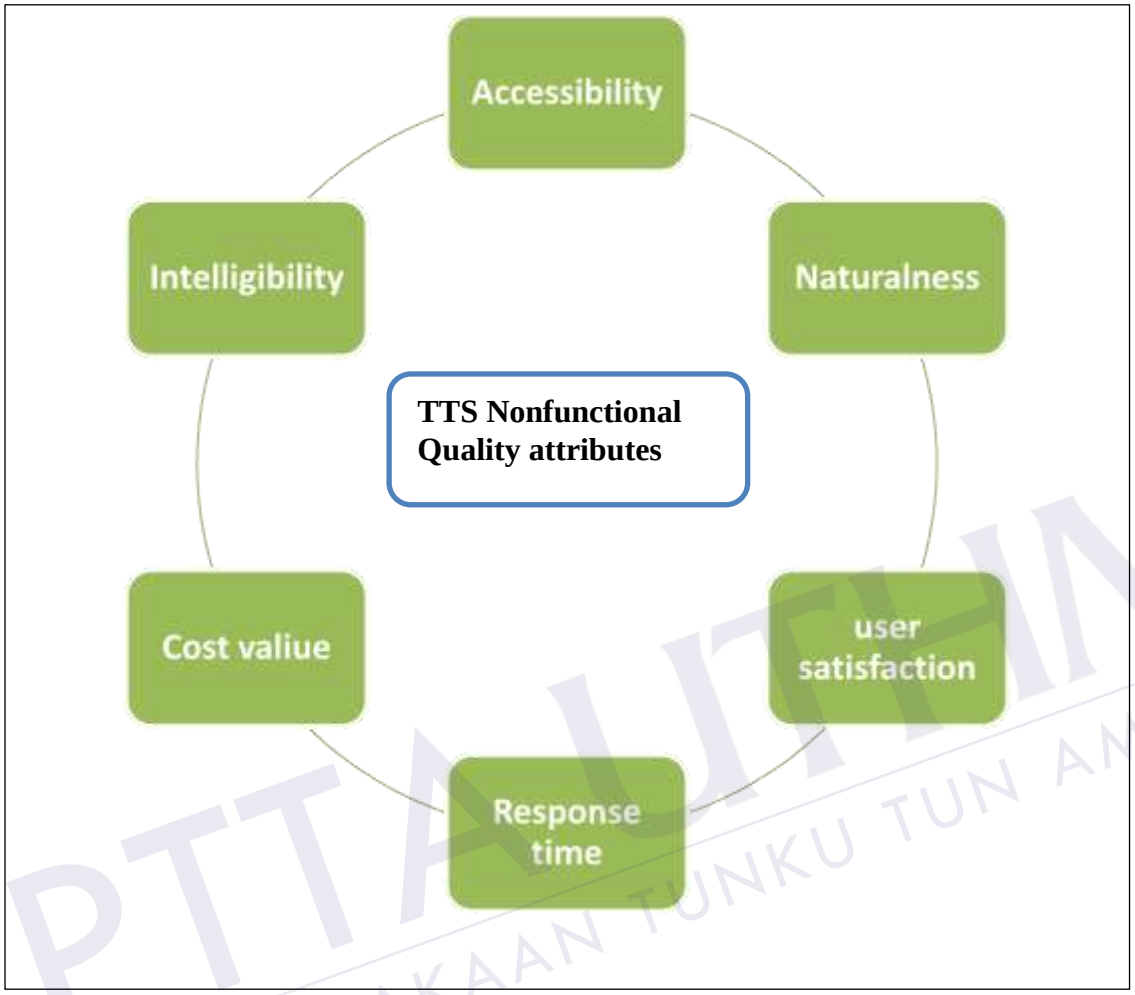

Figure 2.2 Text to speech (TTS) nonfunctional Quality attributes [23].

2.4 Determination of offered QoS

The service developer consider users requirement along with other internal and external factors to determine the optimum practical service quality levels. This framework shows how quality may be determined.

REFERENCES

[1] M. Patil, R. S. Kawitkar, 'Syllable ' Concatenation for Text to Speech Synthesis for Devnagari Script," Int. J. Adv. Res. Eng. Comput. Sci. Softw., vol. 2, no. 9, pp. 180–184, 2012.

[2] K. Manivannan , Ravichandran, "A Framework to Enhance Quality of Service for Content Delivery Network Using Web Services: A Review". IJCSI. Vol8 issue5. pp 275-282. 2011.

[3] D. Sasirekha, E. Chandra, "Text to Speech: A simple Tutorial". International Journal of Soft Computing and Engineering, 275-278, 2012.

[4] W3C− Composite Capability/ Preference Profiles(CC/PP) http://www.w3.org/Mobile.

[5] P. Mohapatra, H. Chen, "A Framework for Managing QoS and Improving Performance of Dynamic Web Content" East Lansing, MI, USA.

[6] S. Moeller, Y. C. Wai, N Cote, TFalk, A Raake, & A Waltermann, "Speech Quality Estimation: Models and Trends". IEEE Signal Processing Magazine. 18-28. 2011.

[7] Y. Y. Chan, "Evaluation of TTS Systems in Intelligibility and Comprehension". Association for Computational Linguistics. PA USA. 153-165. 2011.

[8] M. F. Md Fudzee and J. Abawajy, "Multimedia Content Adaptation Service Discovery Mechanism: Multimedia Tools Applications". Springer Science. New York. 2013.

[9] J. S. Kallimani, K. G. Srinivasa, and E. R. B, "Normalization of Non Standard Words for Kannada Speech Synthesis," vol. 1, no. 1, pp. 21–26, 2012.

[10] H. Bourlard, J. Dines, M. Magimai-doss, P. N. Garner, D. Imseng, P. Motlicek, H. U. I. Liang, L. Saheer, and F. Valente, “Current trends in multilingual speech processing,” vol. 36, no. October, pp. 885–915, 2011.

[11] CH. Wu, Y Huang, “Natural speech synthesis based on hybrid approach” IEEE international conference. pp. 250–254, 2014.

[12] Gueye, Bassirou, et al. "QoS4WSC: A Framework for Web Services Composition based on QoS constraints." International Journal of Computer Information Systems and Industrial Management Applications, 2013.

[13] Kastner, Margit, and Brigitte Stangl. "Exploring a Text-to-Speech Feature by Describing Learning Experience, Enjoyment, Learning Styles, and Values--A Basis for Future Studies." System Sciences (HICSS), 2013 46th Hawaii International Conference. IEEE, 2013.

[14] M. F. Md Fudzee and J. Abawajy. “Request-Driven Cross-Media Content Adaptation Technique,” in Developing Advanced Web Services through P2P Computing and Autonomous Agents: Trends and Innovations, K. Ragab, T. Helmy, and A. E. Hassanien, Eds. IGI Global, ch. 6, pp. 91–113, 2010.

[15] Kastner, Margit, and B Stangl. "Exploring a Text-to-Speech Feature by Describing Learning Experience, Enjoyment, Learning Styles, and Values--A Basis for Future Studies." System Sciences (HICSS), 46th Hawaii International Conference on. IEEE, 2013.

[16] K. Kondo, “Estimation of speech intelligibility using objective measures,” Appl. Acoust., vol. 74, no. 1, pp. 63–70, Jan. 2013.

[17] M. Kastner and B. Stangl, “Exploring a Text-to-Speech Feature by Describing Learning Experience, Enjoyment, Learning Styles, and Values -- A Basis for Future Studies,” 46th Hawaii Int. Conf. Syst. Sci., pp. 3–12, Jan. 2013.

[18] Hifny, Yasser, and M. Rashwan. “Duration modeling for Arabic text to speech synthesis”. INTERSPEECH. 2012.

[19] Schroeder, J. Paul. "The Effects of Age on Processing and Storage in Working Memory Span Tasks and Reading Comprehension." Experimental aging research 308-331, 2014.

[20] Herman, Rosalind, Penny Roy, and F. E. Kyle. "Reading, Dyslexia and Oral Deaf Children: From Research to Practice." 2014.

[21] Papadopoulos, Konstantinos, and Eleni Koustriava. "Comprehension of Synthetic and Natural Speech: Differences among Sighted and Visually Impaired Young Adults." Enabling Access for Persons with Visual Impairment. 147, 2015.

[22] Agboma, Florence, and Antonio Liotta. "Quality of experience management in mobile content delivery systems." Telecommunication Systems. 85-98, 2012.

[23] Bigham, P. Jeffrey, et al. "VizWiz: nearly real-time answers to visual questions." Proceedings of the 23nd annual ACM symposium on User interface software and technology. ACM, 2010.

[24] Bernard, E. David. "Multimodal natural language query system for processing and analyzing voice and proximity-based queries." U.S. Patent No. 7,873,654. 2011.

[25] Couper, P. Mick, et al. "Using Text-to-speech (TTS) for Audio Computer-assisted Self-interviewing (ACASI)." Field Methods, 2014.

[26] Cooke, Martin, et al. "Evaluating the intelligibility benefit of speech modifications in known noise conditions." Speech Communication 55.4. 2013.

[27] Nguyen, Le Thu, R. Harris, and J. Jusak. "Analysis of networking and Application Layer derived metrics for web Quality of Experience." Consumer Communications and Networking Conference (CCNC), IEEE, 2012.

[28] Khan, Asiya, Lingfen Sun, and E. Ifeachor. "QoE prediction model and its application in video quality adaptation over UMTS networks." Multimedia, IEEE, 2012.

[29] Streijl, C. Robert, S. Winkler, and D. S. Hands. "Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives." Multimedia Systems, 2014.

[30] Egger, Sebastian, et al. "Waiting times in quality of experience for web based services." Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on. IEEE, 2012.

[31] L. Wang, et al. "Evaluating text-to-speech intelligibility using template constrained generalized posterior probability." U.S. Patent Application, 2012.

[32] U. Remes, K. Reima, and Mikko. "Objective evaluation measures for speaker-adaptive HMM-TTS systems." Proc. 8th ISCA Speech Synthesis Workshop. 2013.

[33] L. Rabiner, and R. W. Schafer. "Digital Speech Processing." The Froehlich/Kent Encyclopedia of Telecommunications. 237-258, 2011.

[34] Schatz, Raimund, and Tobias Hossfeld. "Web QoE Lecture1: Quality of Experience." Phd School Krakov, Cost TMA 2012.

[35] K. Laghari, et al. "Auditory bcis for visually impaired users: Should developers worry about the quality of text-to-speech readers." International BCI Meeting. 2013.

[36] Waidyanatha, Nuwan, et al. "Mean Opinion Score performance in classifying voice-enabled emergency communication systems." Computer & Information Science (ICCIS), International Conference on. Vol. 2. IEEE, 2012.

[37] S. Arndt, et al. "The effects of text-to-speech system quality on emotional states and frontal alpha band power." Neural Engineering (NER), 6th International IEEE/EMBS Conference on. IEEE, 2013.

[38] Zeiler, D. Matthew, et al. "On rectified linear units for speech processing." Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on. IEEE, 2013.

[39] Latorre, Javier, et al. "Speech intonation for TTS: Study on evaluation methodology." Fifteenth Annual Conference of the International Speech Communication Association. 2014.

[40] M. F. Md Fudzee and J. Abawajy. . "QoS-based adaptation service selection broker." Future Generation Computer Systems 27.3. 2011.

[41] M. F. Md Fudzee and J. Abawajy. "A protocol for discovering content adaptation services." Algorithms and Architectures for Parallel Processing. Springer Berlin Heidelberg, 235-244. 2011.

[42] Eyben, Florian, et al. "Unsupervised clustering of emotion and voice styles for expressive TTS." Acoustics, Speech and Signal Processing (ICASSP), International Conference on IEEE, 2012.

[43] M. F. Md Fudzee and J. Abawajy. . "QoS-based adaptation service selection broker." *Future Generation Computer Systems* 27.3. 2011.

[44] Arndt, Sebastian, et al. "Subjective quality ratings and physiological correlates of synthesized speech." Quality of Multimedia Experience (QoMEX), Fifth International Workshop on. IEEE, 2013.